ELSEVIER

Journal of Pharmaceutical and Biomedical Analysis
16 (1998) 1117–1124

JOURNAL OF
PHARMACEUTICAL
AND BIOMEDICAL
ANALYSIS

Bioanalytical calibration curves: proposal for statistical criteria[1]

Ebi Kalahi Kimanani *

Department of Biometrics and Pharmacokinetics R and D, Phoenix International Life Sciences, 2305 Cohen Street, Ville St. Laurent, Montréal, QC H4R 2N6, Canada

Received 13 December 1996; received in revised form 6 February 1997

Abstract

Curve fitting procedures for bioanalytical assays are based on classical linear least squares (LSE) theory. A common procedure is to select among various models and weighting factors using the R^2 as a goodness-of-fit criterion. It is questionable whether R^2 is the most appropriate criterion for model selection. This is compounded by an often subjective removal of outliers. In this article, statistical curve fitting and diagnostic criteria are proposed. The fitting procedure is a Box-Cox-type power transformation of the data. The optimal transformation is obtained as the one that minimises the sum of squared deviations. Potential outlying standards are screened during the diagnostics stage as those whose jackknife percent deviations exceed 20%. The main advantage of this method is that it is objective and uniformly applicable across analytical techniques. Furthermore, the optimal transformation obtained in this way is unique. The results are demonstrated by comparing the power model to the R^2 approach through the statistical analysis of 2094 analytical batches for 91 projects using various analytical techniques, namely GC, HPLC, LCMS and GCMS. The results indicate that the power model is robust and that QC batch acceptance using the power model is at least as good as the current method. These results hold true across all analytical techniques. It is thus strongly suggested that curve fitting and standard outlier detection for bioanalytical assays should be based on a power model and on jackknife percent deviations method with acceptable cut-off values.

Keywords: Bioanalytical assays; Calibration curves; Power transformation; Jackknife percent deviation; Simulations; Standard curves

1. Introduction

The concentration of a drug present in a biological matrix sample is determined using extraction methods in analytical chemistry to obtain a response value and then estimating the concentration as the abscissa value corresponding to the response on a calibration curve. This curve has concentration (denoted by X) on its abscissa and the response (Y) (either peak height ratio or peak area ratio) on the ordinate axis. The concentrations and responses used to fit the calibration curve are referred to as standards as they are

* Tel: +1 514 3330033; fax: +1 514 3337666.

[1] Presented at the Analysis and Pharmaceutical Quality Section of the Eleventh Annual American Association of Pharmaceutical Scientists Meeting, October 1996, Seattle, Washington, USA.

0731-7085/98/$19.00
PII S0731-7085(97)00064-2

derived from matrices with known concentrations. In this article, the current procedure to select the best calibration curve for a given set of standards is discussed and a new method proposed.

Any curve-selection criteria involves evaluating the appropriateness of a given curve by comparing its goodness-of-fit relative to the rest of the models under consideration. The simplest and most commonly used goodness-of-fit criterion is the multiple correlation coefficient, also called the coefficient of determination, denoted by R^2. The handiness of this measure may overshadow its real value which is as an indicator of the proportion of variability in the response explained by the linear regression. Its blind use could therefore lead to wrong deductions about the functional relationship between response and concentration. Other properties of R^2 serve to caution its use, namely, that R^2 increases with additional variables irrespective of their predictability value. Further, since it depends on the range of X, a large R^2 does not imply a steep slope, neither does it measure the appropriateness of the linear model [1,2] and hence should not be used to compare models if there are transformations on the Y variable. Another problem with the current procedure is that the search for the optimal curve is limited to a small number and it is often difficult to justify the use of the chosen algorithm for negligible differences in R^2. This is illustrated in Fig. 1 which shows standard data points obtained from a plasma sample of Selegeline in LCMS.

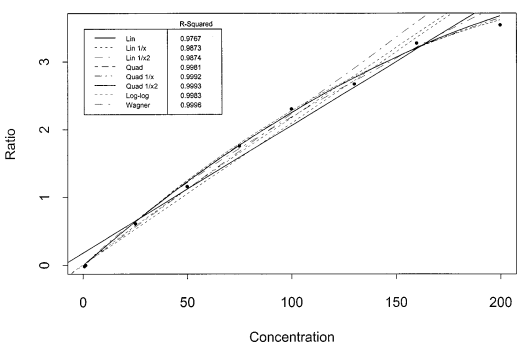

Fig. 1. An illustration of the ambiguity of R^2 on a selection criteria of a regression model from the currently used set of 8 regressions.

Superimposed on this data are the 8 typically used regression types. As can be seen in this plot, all the regressions seem to fit equally well (in terms of R^2). However, it is not clear from the plot which regression to use, neither does the R^2 help since all the values are greater than 0.97 with negligible differences between quite a few of them. Most calibration curves using the current method have, to some extent, this ambiguity of regression type. This problem is further stressed by drug regulatory agencies who require that the chosen algorithm be justified if different from the simplest and, especially, that weighting factors be justified [3].

In this paper an objective procedure for selecting transformations that addresses the above issues is presented. This procedure is based on the class of Box-Cox power transformations using the minimum sum of squared deviations to choose the optimal transformation. The proposed procedure is compared to the current method through an analysis of over 2000 datasets.

2. Power transformation procedure

2.1. Background

Data transformation prior to performing a regression analysis is a common practice in statistical analysis. Indeed, some statistical software packages include some form of routine data transformation in an exploratory/data-descriptive module. The most common objective for transformation is usually to meet the requirements of classical linear regression, e.g. to stabilize the error variance and normalize the error distribution [4,5]. In bioanalytical assays, it is believed that the variance of the ratio of peak heights increases with increasing drug concentration. However, the variance weighting function is not known. The purpose of this study is to present an objective statistically valid regression fitting procedure that implicitly fits the proper variance weighting function.

Given a number of standards, n (usually 8 or 10) with responses $Y_1, Y_2, \ldots, Y_n$ and concentrations $X_1, X_2, \ldots, X_n$, the typical statistical data transformation problem has three components:

- a model linking the transformed response to the transformed concentration,
- a set of transformations from which to choose,
- a goodness-of-fit function (to discriminate between transformations).

A good example is Breiman and Friedman's [6] alternating conditional expectations (ACE) method. Here, the model equation is given by

$$\theta(y_i) = \phi(x_i) + \epsilon_i$$

for transformations θ and ϕ that may be different. The set of transformations is not fixed but is determined dynamically by iterative alternating conditional expectations. The goodness-of-fit criterion is the scaled residual sum of squares.

The method proposed in this article is a generalized version of the Box-Cox transformations [6] and is similar to ACE in the model equation and also the goodness-of-fit criterion function. In Section 2.2, the power model is presented in greater detail.

2.2. The power transformation

The power transformation is defined as:

$$y^{(\lambda)} = \begin{cases} \dfrac{y^{\lambda} - 1}{\lambda \dot{y}^{\lambda - 1}}, & \text{if } \lambda \neq 0 \\ \dot{y} \log y, & \text{if } \lambda = 0 \end{cases} \quad (1)$$

where $\dot{y}$ is the geometric mean and λ is the power value. The model linking response (Y) and concentration (X) is a simple linear regression between the power transformed X and Y and is given by

$$y_i^{(\lambda)} = \mu + a x_i^{(\lambda)} + \epsilon_i \quad (2)$$

The residual sum of squares (SSR) from the above regression is given by

$$\mathrm{SSR}(\lambda) = \sum (y_i^{(\lambda)} - \mu - a x_i^{(\lambda)})^2 \quad (3)$$

The optimal power transformation λ_{opt} is determined empirically as the value of λ which minimises SSR, the residual sum of squares. Since minimising SSR(λ) is equivalent to maximising the likelihood function [1,5], the power obtained will be doubly optimal in that it will be both the least squares estimate as well as the maximum likelihood estimate of the true power. Once the optimal power is determined, possible quadrature is accommodated by allowing a quadratic term to be added to model Eq. (2). The steps involved in the empirical determination of the optimal power transformation are described below.

2.3. Algorithm A: The power curve fitting algorithm

Step 1: To start, a power search range and precision of the power transformation is selected. The power range is unbounded and could be any set of real numbers. The power precision is also set at this stage, usually linked to the number of iterations.

Step 2: Given a set of concentrations and responses and for each λ in the search range, the following procedure is followed:

- transform the response and concentration according to Eq. (1),
- fit a linear regression between the transformed response and transformed concentration (Eq. (2)),
- obtain SSR(λ) (Eq. (3)).

Step 3: From the whole set of SSR(λ) values, the optimal power transformation, λ is the power value at which SSR(λ) is minimised.

Step 4: Using the optimal power value, a quadratic term in the transformed X is then added to the model and retained if statistically significant at α level 0.10.

This procedure is illustrated in Figs. 2–4, all of which are based on the same data as that in Fig. 1 and were chosen to illustrate the whole power fitting algorithm including the accommodation of significant quadrature. Fig. 2 illustrates fitted regressions for some of the powers in the search range, each with its corresponding SSR. Fig. 3 illustrates a typical profile of SSR as a function of the powers and demonstrates that the optimal value is unique. Fig. 4(a) is an illustration of the linear regression curve corresponding to the optimal power (-0.461) as well as that chosen by the current method (Wagner) and their corresponding

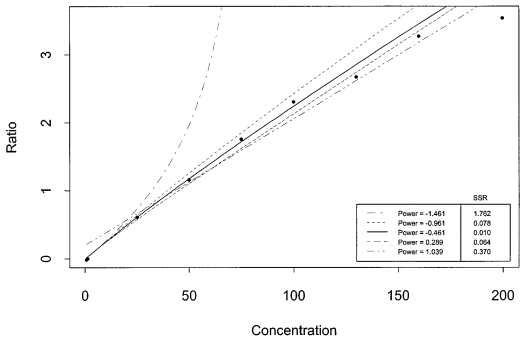

Fig. 2. An example of the curves in the power search range of the power curve finding algorithm and their corresponding scaled residual sum of residuals. The search range is unrestricted. The residuals are scaled in order to be comparable.

R^2. From this curve, the power model seems to miss the apparent curvature in the data. A quadratic term is added and the improved fit is shown in Fig. 4(b). The objectivity of the power method is thus illustrated as well as the uniqueness of choice of the optimal transformation. Once the optimal power is determined, the next step is to screen for potential outlying standards. In Section 2.4, the proposed outlier detection method is described.

2.4. The 20-20 jackknife percent deviation rule

The jackknife percent deviation of a given standard is defined as the percent deviation of the

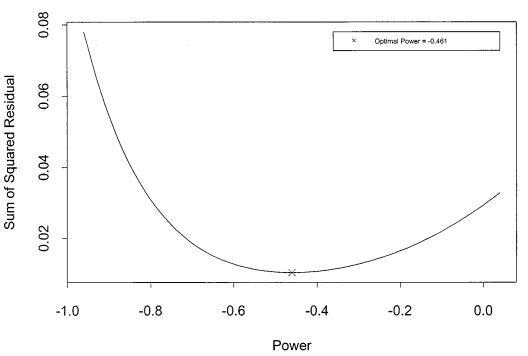

Fig. 3. A typical profile of the sum of searched residuals over the search range. Because of the strict concavity of the function, the optimal power is unique.

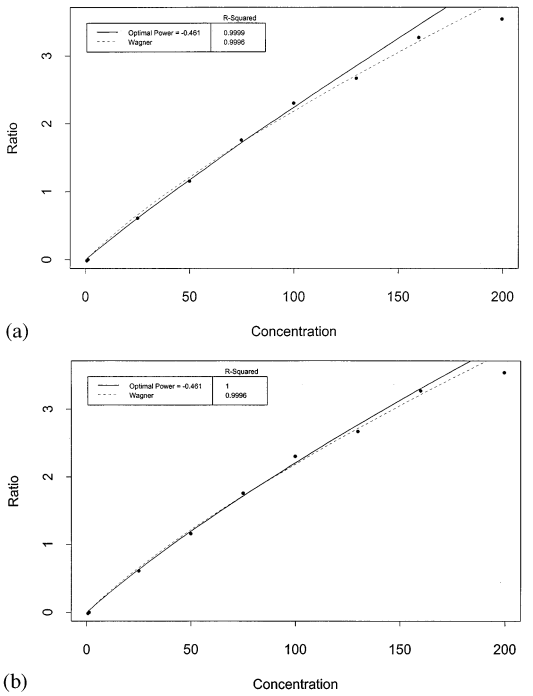

Fig. 4. (a) Comparison of the selected curves using the power model with a simple linear regression and using the current method. (b) Comparison of the selected curves using the power model with a quadratic regression and using the current method.

standard from the regression line that is fit excluding it. Jackknifing is a useful tool in detecting points that may have undue influence in a particular fit but which would not show up as outliers [2]. The proposed jackknife percent deviation rule is defined as follows. To start, the jackknife percent deviation of every standard is calculated. Based on these values, a standard is considered an outlier if it has the highest absolute percent deviation greater than 20%. If an outlier exists and if the total number of outliers is less than 20% of the total number of standards, then the standard is excluded from the analysis and the fitting procedure in algorithm A is repeated. Otherwise, the line is cross-validated using the quality control (QC) samples. Assessing the QC fits is done using the current 4/6: 20-15-10 rule which is described in Appendix A.

Thus the power model consists of four steps: power transformation, linear or quadratic regression, outlier detection and cross-validation using QCs. This model will be compared to the current method through data from real studies as described in Section 3.

3. Data analysis

Given X = concentration and Y = response, the current set of curves from which a calibration curve is chosen are:

Linear: $y = \mu + \beta_1 x + \varepsilon(0, \sigma^2)$

Linear, weighted $1/x$: $y = \mu + \beta_1 x + \varepsilon(0, \sigma^2 x)$

Linear, weighted $1/x^2$: $y = \mu + \beta_1 x + \varepsilon(0, \sigma^2 x^2)$

Quadratic: $y = \mu + \beta_1 x + \beta_2 x^2 + \varepsilon(0, \sigma^2)$

Quadratic, weighted $1/x^2$: y

$= \mu + \beta_1 x + \beta_2 x^2 + \varepsilon(0, \sigma^2 x^2)$

Quadratic, weighted $1/x^2$: y

$= \mu + \beta_1 x + \beta_2 x^2 + \varepsilon(0, \sigma^2 x^2)$

$\text{Log} - \text{log}: \log y = \mu + \beta_1 \log x + \varepsilon(0, \sigma^2)$

Wagner: $\log y = \mu + \beta_1 \log x + \beta_2 (\log x)^2 + \varepsilon(0, \sigma^2)$ (4)

The power model was used to calibrate the standard curves from 2094 analytical batches of data from 91 projects using various analytical techniques namely GC, LC, GCMS, LCMS. The current method of selecting a regression type from the above set according to the maximum R^2 value was also used. The performance of the power model was compared to the current method through concordance tables of run acceptance and outlier rejection rates. The comparison results can be found in Tables 1–3 and in Figs. 5 and 6.

4. Results

4.1. Acceptance rates

All the R^2 values reported have been adjusted for the number of parameters. In all the tables and figures, results from the power model are labelled "power" while those from the current approach are labelled "current". P and L indicate outlier points detected by the power model and the current method, respectively. The overall run acceptance results are shown in Table 1. The results show an overall batch acceptance concordance of around 95%. Concordance is demonstrated by standard data from levonorgestrel in GCMS in which the selected curves from both methods were rejected (Fig. 5) or accepted (Fig. 6) during the QC assessment stage. The 5% discor-

Table 1
Comparison of QC-based run acceptance rates in the power model and in the current method

		Current method		
		Accept	Reject	Total
Power model	Accept	1576	56	1632
	Reject	55	407	462
	Total	1631	463	2094

Concordance = 95%, discordance = 5%.

Table 2
Run acceptance rates comparison stratified by regression type

		Current method		
		Accept	Reject	Total
Linear				
Power model	Accept	1432	52	1484
	Reject	49	367	416
	Total	1481	419	1900
Quadratic				
Power model	Accept	14	0	14
	Reject	1	7	8
	Total	15	7	22
Wagner				
Power model	Accept	130	4	134
	Reject	5	33	38
	Total	135	37	172

Concordance = 95%, discordance = 5%.

Table 3
Run acceptance rates comparison stratified be analytical technique

		Current method		
		Accept	Reject	Total
GC[a]				
Power model	Accept	258	11	269
	Reject	14	109	123
	Total	272	120	392
GCMS[b]				
Power model	Accept	413	21	434
	Reject	18	125	143
	Total	431	146	577
HPLC[c]				
Power model	Accept	576	14	590
	Reject	11	61	72
	Total	587	75	662
LCMS[d]				
Power model	Accept	328	10	338
	Reject	12	112	124
	Total	340	122	462

[a] Concordance = 94%, discordance = 6%; [b] concordance = 93%, discordance = 7%; [c] concordance = 96%, discordance = 4%; [d] concordance = 95%, discordance = 5%.

dance may be due to the imprecision in the current method of curve selection that was demonstrated in the simulation study reported in [6]. It

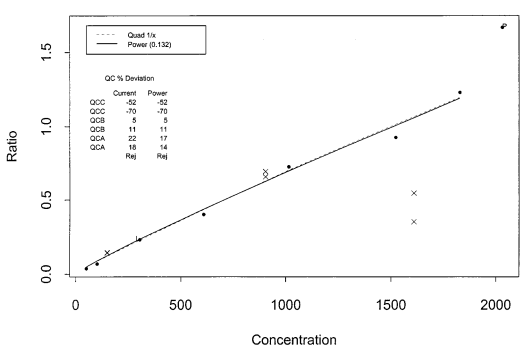

Fig. 5. Concordance between the power model (optimal power = 0.132) and the current method (quadratic weigthed $1/X$) in which both methods rejected the curve. QCA, QCB, QCC correspond to the low, medium and high concentration QC samples respectively. The symbol x marks the QC value on the graph while L and P mark outliers detected by the current method and by the power method, respectively.

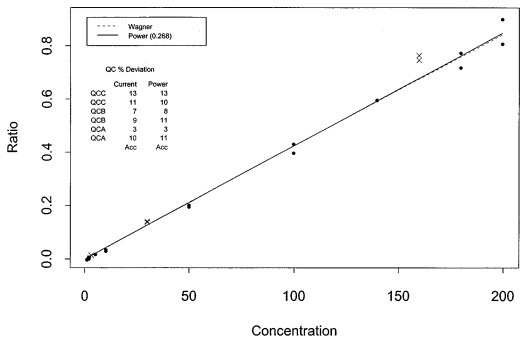

Fig. 6. Concordance between the power model (optimal power = 0.058) and the current method (quadratic weigthed $1/X$) in which both methods accepted the curve. QCA, QCB, QCC correspond to the low, medium and high concentration QC samples respectively. The symbol x marks the QC value on the graph while L and P mark outliers detected by the current method and by the power method, respectively.

may also be due to minor differences in the QC percent deviations as is shown in Fig. 7 (standard data from carboxyterfenadine in GC) and Fig. 8 (equilin in GCMS).

These concordance–discordance values are maintained whether the results are stratified by regression type (Table 2) or analytical technique (Table 3).

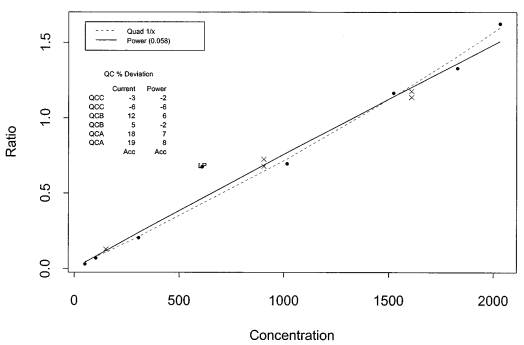

Fig. 7. Discordance due to minor differences in QC percent deviation. QCA, QCB, QCC correspond to the low, medium and high concentration QC samples respectively. The symbol x marks the QC value on the graph while L and P mark outliers detected by the current method and by the power method, respectively.

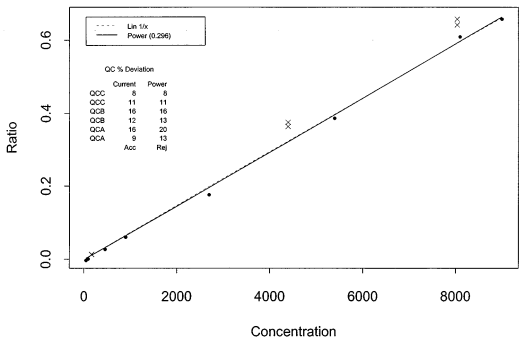

Fig. 8. Discordance due to minor differences in QC percent deviation. QCA, QCB, QCC correspond to the low, medium and high concentration QC samples respectively. The symbol *x* marks the QC value on the graph while *L* and *P* mark outliers detected by the current method and by the power method, respectively.

4.2. Outlier detection

The 20-20 jackknife percent deviation allowed for a slightly lower outlier rate of 5.7% in the power model compared to the current rate of 5.9%. The outlier rate is the proportion of points rejected as outliers in the whole dataset.

5. Discussion

5.1. Power model

The power model discussed in this article was proposed by Box and Cox [5] for the purpose of linearising the mean function and stabilising the error variance. It was intended for data analysis situations such as that presented by the standard curve problem. It generalises the weighting scheme that is currently used to fit calibration curves. Previous simulations [6] and experimental data analysis reported in this article demonstrated additional advantages of the power transformation method over the current method, which are that:

- it is based on an objective choice of transformations;
- the range of transformations from which to choose is continuous and unbounded;
- the power transformation range includes the current set of regression models;
- a unique optimal transformation is obtained, which eliminates the problem of trying to justify the choice between two equally optimal curves, as frequently happens in the current procedure;
- valid quadrature is accommodated by retaining a quadratic term in the model if it improves the fit over the linear either through a statistical significance test or by comparing the back-calculated standards for the two models;
- it is highly specific, as was observed in the simulations;
- it is simple to understand and implement.

The main advantage of the proposed 20-20 jackknife percent deviation rule for outlier detection is that it is objective, statistically valid and is applied uniformly. It resulted in slightly fewer outliers than the current method. Using the power model and the outlier detection rule, results from extensive data analysis demonstrated a 95% concordance between the power model and the current procedure. The discordance may be due to a 5% error rate resulting from a reduced precision in curve selection using the current procedure.

5.2. Conclusion

In general, regression fitting is considered a solved statistical problem. The power model presented in this article is one of the methods that has been suggested by statisticians for fitting the correct variance weight. It is well known and widely used by data analysts as is the jackknife procedure for outliers. Simulation results showed that the power model is highly specific in that the optimal power thus obtained is an accurate estimate of the underlying true variance weighting factor. Experimental data analysis results showed that using the proposed criteria, calibration curve fitting can be made more scientifically valid and objective without increasing production costs (the latter observed from the 93% concordance between the proposed criteria and the current procedure). On the strength of these results, it is recommended that: (1) the power model be used as a scientifically valid procedure for fitting cali-

bration curves; and (2) the jackknife percent deviation with acceptable cut-off values be used for outlier detection.

Acknowledgements

This research was motivated by John Hooper and Richard Lalonde. The author would also like to thank Jean Lavigne, Tony Chilton, Daniele Ouellet and Diane Potvin for critical reviews of the manuscript. The project would not have been completed without the full collaboration of the bioanalytical departments at Phoenix International Life Sciences.

Appendix A. Cross validation

A.1. QC assessment rule

This rule requires that the quality control samples at three concentration levels (low, medium and high) be assayed in duplicate in each run. For the run to be accepted, at least 4 out of the 6 samples should have estimated concentrations that are within 20% (for the low concentration level), 15% (medium level) and 10% (high level) of their known values and that at least one of the two values at each concentration level must be within these limits. This rule is criticized in [7,8].

A.2. Back calculation of concentrations

In order to back calculate the concentration, i.e. estimate the corresponding concentration given a response, note that in a simple linear regression, the prediction of a new value depends on the data implicitly through the parameters. Thus, given (X_1, Y_1), (X_2, Y_2),…, (X_n, Y_n), and the fitted regression

$$Y_i = \mu + \beta X_i \tag{A1}$$

then the back calculated new concentration X_{new} given a new response Y_{new} is

$$X_{new}/\{(X_1, Y_1), (X_2, Y_2),\ldots, (X_n, Y_n), Y_{new}\} = (Y_{new} - \mu)/\beta \tag{A2}$$

Using this logic, the back calculation in the power transformation case of a new value given a new response is given by

$$\theta(X_{new})/\{(X_1, Y_1), (X_2, Y_2),\ldots, (X_n, Y_n), Y_{new}\} = (\theta(Y_{new}) - \mu)/\beta \tag{7}$$

where

$$\theta(Y_{new}) = \begin{cases} \dfrac{Y_{new}^{(\lambda_{opt})} - 1}{\lambda\left(\prod_{i=1}^{n} Y_i\right)^{(\lambda-1)/n}}, & \text{if } \lambda_{opt} \neq 0 \\ \left(\prod_{i=1}^{n} Y_i\right)^{1/n} \log Y_{new}, & \text{if } \lambda_{opt} = 0 \end{cases}$$

References

[1] D.C. Montgomery, E.A. Peck, Introduction to Linear Regression Analysis, Wiley, New York, 1992.
[2] N. Draper, H. Smith, Applied Regression Analysis, Wiley, New York, 1981.
[3] J. Hooper, Bioanalytical validation: a North American view, in: H.H. Blume and K.K. Midha (Eds.), Bio-International 2: Bioavailability, Bioequivalence and Pharmacokinetics Studies, Medpharm Scientific Publishers, Stuttgart, 1995 .
[4] M.S. Bartlett, The use of transformations, Biometrics 3 (1947) 39–52.
[5] G.E.P. Box, D.R. Cox, An analysis of transformations, J. R. Stat. Soc. 26 (1964) 211–252.
[6] L. Breiman, J.H. Friedman, Estimating optimal iransformations for multiple regression and correlation (with discussion), J. Am. Stat. Assoc. 80 (1985) 580–619.
[7] E.K. Kimanani, J. Lavigne, Bioanalytical calibration curves: Variability of optimal powers between and within analytical methods, J. Pharm. Biomed. 16 (1998) 1107–1115.
[8] R.O. Kringle, An assessment of the 4-6-20 rule for acceptance of analytical runs in bioavailability, bioequivalence and pharamacokinetic studies, Pharm. Res. 11 (4) (1994) 556–560.

.